Facial expression analysis and the affect space

Varun Jain, Evanthia Mavridou, James L. Crowley, Augustin Lux

Facial Expression Analysis and the Affect Space

Varun Jain, Evanthia Mavridou, James L. Crowley, Augustin Lux
INRIA Grenoble Rhône-Alpes Research Center
{varun.jain, evanthia.mavridou, james.crowley, augustin.lux}@inria.fr

***Abstract*—In this paper we present a technique for facial expression analysis and representing the underlying emotions in the affect space. We develop a purely appearance based approach using Multi-scale Gaussian derivatives and Support Vector Machines. The technique is validated on two different databases. The system is shown to generalize well and performs better than the baseline method.**

***Index Terms*—Automated Facial Expression Analysis, Affect Recognition, Multi-scale Gaussian Derivatives**

I. Introduction

Facial expressions are a mirror to human emotions and an important component of human to human interaction. Human computer interaction requires the same ability to read emotions from facial expressions. Ekman introduced the concept of six basic emotions that are universally recognizable [1]. In [2] he presented the Facial Action Coding System(FACS), a taxonomy to describe facial expressions in terms of individual muscle movements.

FACS based approaches have been adopted in a variety of vision systems such as the Computer Expression Recognition Toolbox(CERT) [3]. Such systems are trained to estimate the Action Unit(AU) intensities which can then be used to assign one of the six basic emotion labels to that image or frame. The problem arises when the expression in the image is not associated with any of the six basic emotions.

An alternative to such a structured approach is to represent the underlying emotions in a multidimensional emotion space. Shin in [4] uses component analysis techniques to recognize emotions and map them to the affect space. Another method was presented by Dahmane and Meunier in [5]. The authors used Gabor wavelets and Support Vector Machines on the Semaine database [6] and they use 4 dimensions (Activation, Expectation, Power and Valence) to represent the emotions that underlie the facial expressions.

In [7] the authors argue that three dimensions are enough to represent any emotion. In this paper we use the affect space model developed by Russell and Mehrabian and compare our results for Pleasure and Arousal with the results from the technique presented in [5].

Two common ways to describe image features are: appearance based methods and geometric feature based methods. The latter involves detection and tracking of facial keypoints such as the lip corners, nostrils and eyes. This detection and tracking is done with the help of computationally expensive vision techniques and are not very robust.

The approach we present here does not involve identification of any landmarks on the face and just like the appearance based technique discussed in [5], the image filters are applied to the whole-face to obtain the feature vector.

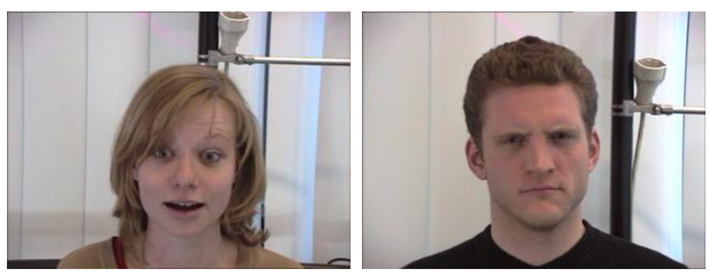
II. Affect Space and datasets used

Russell and Mehrabian in [7] describe a 3-dimensional affect space(Pleasure/Displeasure, Arousal/Sleep, Dominance/Submissiveness) model that can be used to describe the emotional state of a person. Experiments support that these 3 dimensions are sufficient to represent all human emotions.

The Pleasure/Displeasure Scale measures the pleasantness of an emotion while the Arousal/Sleep Scale measures the intensity of the emotion and the Dominance/Submissiveness Scale represents the controlling and dominant nature of the emotion. The third axis of Dominance remains controversial and there is evidence to suggest that there is a high correlation between dominance and the other two axes[8].

Our approach was tested on the Cohn-Kanade [9] and FEED [10] datasets. The FEED dataset was collected at the Technical University of Munich. The dataset was generated as a part of the European Union FG-NET project [11].

The FEED dataset does not contain posed emotions, the emotions were elicited by showing video clips to the participants. The database contains images from 18 individuals for 6 basic emotions along with the neutral face.



(a) Surprised (b) Angry

Fig. 1: Example Images from the FEED dataset

We map the basic emotions to the Pleasure-Arousal space as shown in table 1 in accordance with the Pleasure-Arousal values provided by Mehrabian. Instead of using numerical values we assign class labels (+P -P, +A -A) to perform binary classification.

Emotion	P Label	A Label
Joy	+	+
Sadness	-	-
Surprise	+	+
Anger	-	+
Disgust	-	+
Fear	-	+

TABLE I: Labels for the 6 basic emotions

The Cohn-Kanade and FEED databases were re-annotated with these class labels. The Cohn-Kanade database was used for training and validation while the FEED database was used for testing.

III. Multi-scale Gaussian Derivatives

Gaussian derivatives can efficiently describe the neighborhood appearance of a pixel for pattern recognition tasks [12]. For images this is done by calculating different orders of Gaussian derivatives normalized in scale and orientation at every pixel.
The basic Gaussian function is defined as:

$$G(x, y; \sigma) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

Here σ is the scale factor or variance and defines the spatial support. This function measures the intensity of the neighborhood and does not contribute to the identification of the neighborhood and can be omitted.
First order derivatives provide information about the gradient (intensity and direction) whereas the second order derivatives provide the information about image features such as bars, blobs and corners. Higher order derivatives are only useful if the second order derivatives are strong otherwise they just contain image noise.
Obtaining scale invariant features is not a trivial task. Several methods have come up in the past addressing this problem. Lindeberg in [13] suggests that Gaussian derivatives be calculated across scales to get scale invariant features and then Lowe in [14] defines the intrinsic or characteristic scale as the value of the scale parameter at which the Laplacian provides a local maximum. The computational cost of directly searching the scale axis for this characteristic scale can be prohibitively expensive. A cost-effective method for computing Multi-scale Gaussian derivatives has been discussed in detail in [15]. The next section is about Principal Component Analysis (PCA) and why we need it.

IV. Principal Component Analysis

The region of the image containing the face is normalized to 64 X 64 pixels, this particular size is chosen after extensive experimentation where normalized images of 64 X 64 pixels gave the best accuracy. We calculate several orders of derivatives at 2 levels of scale for every pixel but it leads to an enormous feature vector. Therefore we divide the image into cells of 4 X 4 pixels and the feature vector contains the mean and standard deviation of the descriptor values (gaussian derivatives) for each cell of 4 X 4 pixels.

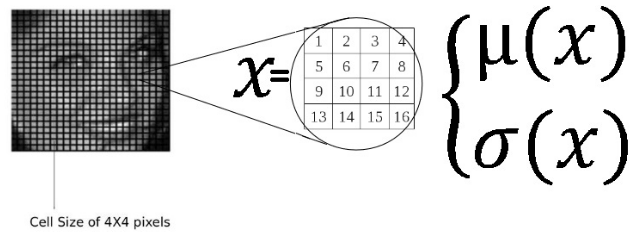


Fig. 2: The image divided into cells of 4 X 4 pixels.

Principal Component Analysis is used for dimensionality reduction which reduces the prediction time when the Support Vector Machines are used for classification. Correlation in the data can be reduced by transforming the original dimensions into new dimensions which are a linear sum of the original dimensions but are linearly uncorrelated. Then these new dimensions are ranked according to the variance i.e. the dimension which accounts for the most variability in the data gets the first rank and so on [16].
PCA is done by eigenvalue decomposition of the data correlation matrix after normalizing the data for each dimension. PCA provides us with scores and loadings. The scores are the transformed values corresponding to the data point and loadings are the coefficients the original variable should be multiplied with to get the score.

V. Support Vector Machines

Support Vector Machines (SVM) belong to a family of non-probabilistic linear classifiers [17]. The Radial Basis kernel provides the best accuracy for the particular application and is represented the following equation:

$$K(x_i, x_j) = e^{-\frac{||x_i - x_j||^2}{2\sigma^2}} \tag{2}$$

We use a soft margin SVM. Soft margin SVMs are used when the classes are not separable even after transforming the data to a higher dimension. The condition for the optimal hyper-plane can be relaxed by including an extra term ξ [18]:

$$y_i(X_i^T W + b) \geq 1 - \xi_i, (i = 1, ..., m) \tag{3}$$

For minimum error,ξ_i should be minimized as well as $||W||$, and the objective function becomes:

$$\begin{aligned} &minimize \quad W^T W + C \sum_{i=1}^{m} \xi_i^k \\ &subject\,to \;\; y_i(X_i^T W + b) \geq 1 - \xi_i, \;\; and \;\; \xi_i \geq 0; (i = 1, ..., m) \end{aligned} \tag{4}$$

Here C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the training error. $1/\gamma$ or σ is the width of the radial basis kernel. The C-penalty parameter is chosen using cross validation. For the data in hand C = 15 and $\sigma = 280$ lead to the highest classification accuracy for Arousal and C = 10 and $\sigma = 190$ are found to be the optimum parameters for Pleasure.

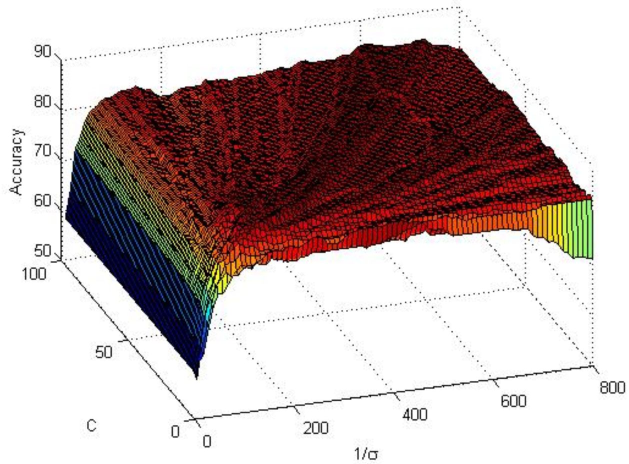


Fig. 3: Graph of Classification Accuracy vs. C-parameter and σ for Pleasure.

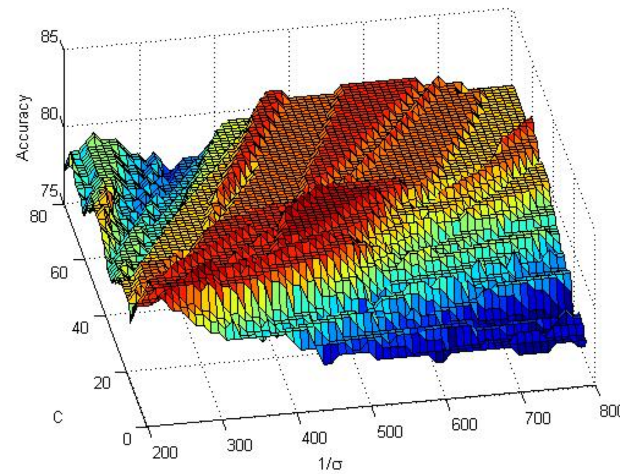


Fig. 5: Graph of Classification Accuracy vs. C-parameter and σ for Arousal.

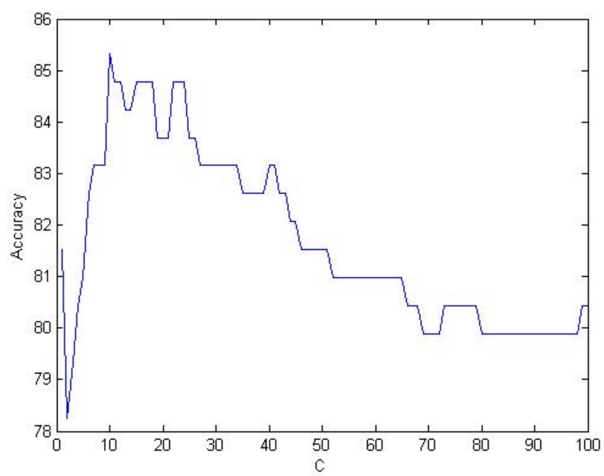


Fig. 4: Graph of Classification Accuracy vs. C-parameter at $\sigma = 81$ for Pleasure.

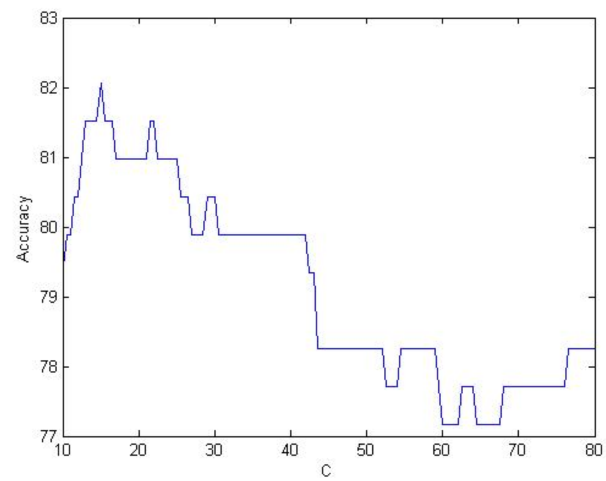


Fig. 6: Graph of Classification Accuracy vs. C-parameter at $\sigma = 81$ for Arousal.

VI. The Approach

Face detection is performed on the images in the dataset using the OpenCV face detector [19]. Following that a half-octave gaussian pyramid is constructed over a normalized imagette of the face. This is followed by dimensionality reduction by PCA and regression using Support Vector Machines. The figure below illustrates the process.

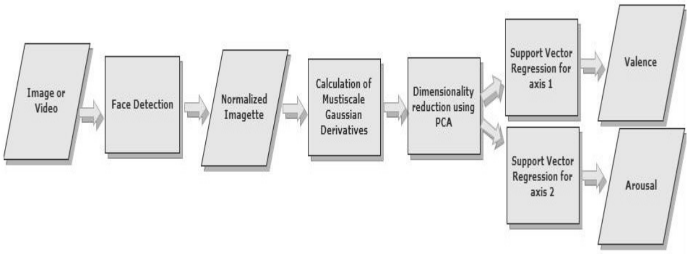


Fig. 7: Schematic of our approach.

VII. Results

We divide the Cohn-Kanade database into two, 70 percent of the images are used for training and the rest for validation. The database is split several times and the accuracy is calculated for every split and the average is calculated. The ROC for the two SVM's used are shown in the figures below. The first ROC is for the SVM trained for detecting Pleasure and the second one for Arousal.

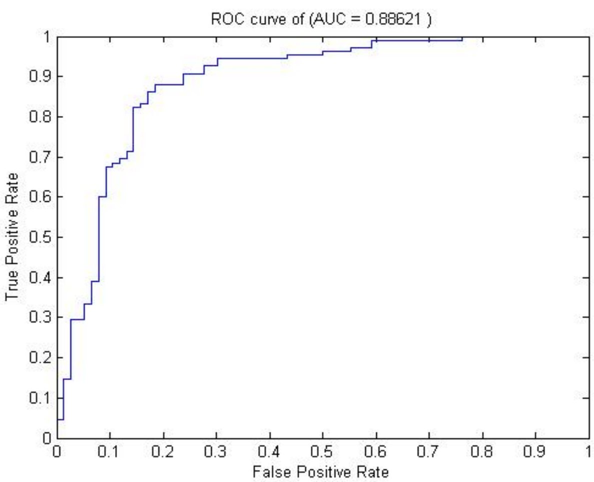


Fig. 8: ROC of the classifier for Pleasure

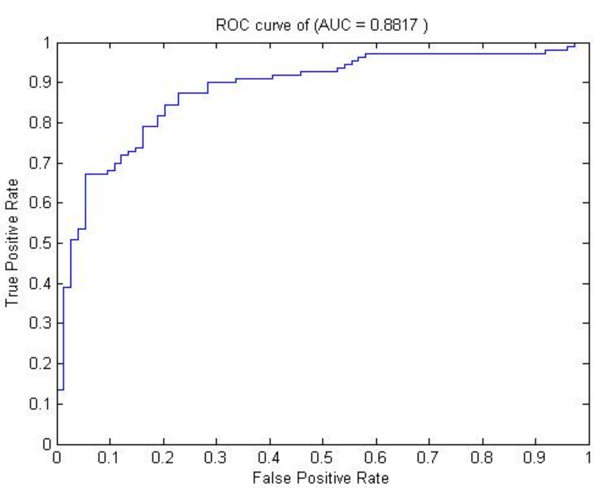


Fig. 9: ROC of the classifier for Arousal

The accuracy of our approach over the Cohn-Kanade set is 85.32,82.06 percent for pleasure and arousal respectively. On the other hand the approach developed by Dahmane and Meunier achieves an accuracy of only 71.80,74.94 percent for pleasure and arousal respectively.

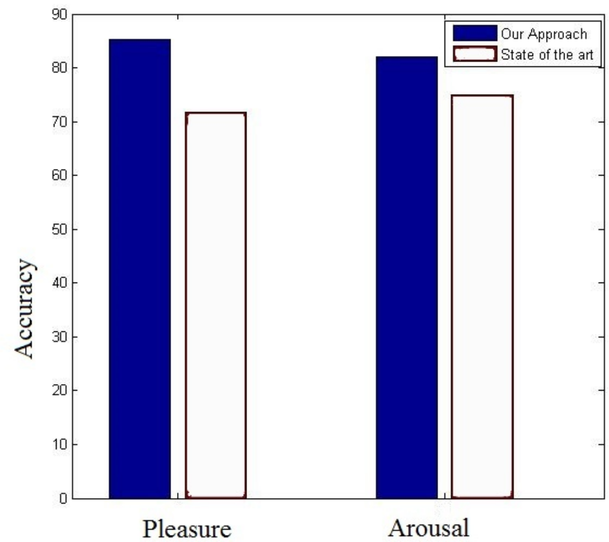


Fig. 10: Comparison of results

We also see that it takes much less time to computer Gaussian derivatives using the half-octave pyramid as compared to Gabor features because of the ineteger coefficient Half-Octave Pyramid used. The table below shows the time to calculate the features for the complete Cohn-Kanade database using the two techniques on the same machine(Intel Xeon Quad-Core 3GHz, 4GB RAM).

	Multi-scale Gaussian Derivatives	Gabor Energy Filters
Calculation Time(sec)	5.36	20.37

TABLE II: Comparison of time required for calculating the two types of features

PCA reduces the prediction time by a factor of over 60, table 3 compares the prediction time with and without using PCA.

	SVM with PCA	SVM without PCA
Prediction time(sec)	0.0155	0.8495

TABLE III: Comparison of prediction time with and without using PCA

Table 4 shows the prediction time of our technique versus the state of the art because our feature vector is much smaller.

	Our Approach	State of the art
Prediction Time(sec)	0.0155	1.06

TABLE IV: Comparison of prediction time

Our approach is then tested on the FEED database and the accuracy for Pleasure-Displeasure is 70.73% while it is 70.08% for Arousal-Nonarousal.

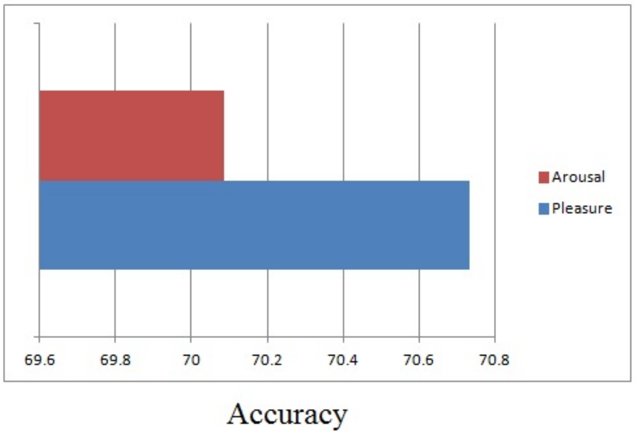


Fig. 11: Results on the FEED database

VIII. Conclusion

We have presented a novel method to analyze facial expressions and represent the underlying emotion in the affect space. Not only is our performance better than that of the baseline approach, it is also faster at descriptor calculation and prediction. The approach performs better than the benchmark technique and is easily adaptable to mobile systems.
Codes exist for calculating Multi-scale Gaussian derivatives on embedded systems using only integer coefficients.

References

[1] Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200, 1992.

[2] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, 1978.

[3] G. Littlewort, J. Whitehill, T. Wuand I. Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (cert). In *Proceedings of the 2011 IEEE International Conference on Automatic Face and Gesture Recognition*, pages 298–305, 2011.

[4] Young-suk Shin. Recognizing facial expressions with pca and ica onto dimension of the emotion. In *Proceedings of the 2006 Joint IAPR International Conference on Structural, Syntactic, and Statistical Pattern Recognition*, SSPR'06/SPR'06, pages 916–922, Berlin, Heidelberg, 2006. Springer-Verlag.

[5] Mohamed Dahmane and Jean Meunier. The computer expression recognition toolbox (cert). In *Affective Computing and Intelligent Interaction 2011*, pages 351–358, 2011.

[6] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17, 2012.

[7] James A. Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977.

[8] Christian Becker-Asano. *WASABI:Affect Simulation for Agents with Believable Interactivity*. PhD thesis, University of Bielefeld, 2008.

[9] J. Cohn T. Kanade and Y.-L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, March 2000.

[10] Frank Wallhoff. Facial expressions and emotion database. http://www.mmk.ei.tum.de/w̃af/fgnet/feedtum.html, Technical University of Munich, 2006.

[11] http://www prima.inrialpes.fr/FGnet/. Face and gesture recognition working group.

[12] Varun Jain and James L. Crowley. Smile detection using multi-scale gaussian derivatives. In *Proceedings of the 12th WSEAS International Conference on Signal Processing, Robotics and Automation*, pages 149–154, February 2013.

[13] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Press, 1994.

[14] David G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, 1999.

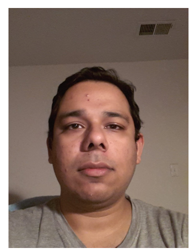
[15] J.L. Crowley and O. Riff. *Springer Lecture Notes on Computer Science*, volume 2695, chapter Fast Computation of Scale Normalized Gaussian Receptive Fields, pages 584–598. 2003.

[16] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 2002.

[17] V.N. Vapnik. *Statistical Learning Theory*. Wiley, NY, 1998.

[18] C. Cortes and V.N. Vapnik. *Machine Learning*, volume 20, chapter Support-Vector Networks, pages 273–297. Springer, 1995.

[19] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 20(17):137–154, 2004.



Varun Jain Varun Jain has a Bachelors degree in Electronics Engineering and a Masters in System Engineering. He has been pursuing his PhD at the INRIA Grenoble Rhône-Alpes Research Center since October 2011. He works on facial expression analysis and Affective computing. His areas of interest include Image Processing, Machine Learning and Human-Computer Interaction.

Evanthia Mavridou Evanthia Mavridou received her BA in applied informatics at the University of Macedonia, Greece, in 2007. During her BA studies, she performed one semester in 2005 at the University of Joensuu, Finland. From 2007 to 2010, she worked as a web programmer in Thessaloniki, Greece. In 2011, she received the MS in image, computer vision and robotics at the University of Crete, Greece. In 2011, she has started her research career in computer vision at the INRIA Grenoble Rhône-Alpes research center, France, where she is currently pursuing her PhD under the supervision of Prof. James L.Crowley. She is working on the exploration and development of efficient compact and robust image description methods. Her research interests include image description, feature detection and recognition.

James L. Crowley James L. Crowley directs the PRIMA Research Project group of the LIG laboratory at the INRIA Grenoble research center in Montbonnot, France. He holds the post of Professor at the Institut National Polytechnique de Grenoble (INPG), where he teaches courses in Computer Vision, Signal Processing, Pattern Recognition and Artificial Intelligence at l'ENSIMAG (Ecole National Superieure d'Informatique et de Mathematiques Appliques). From 2003 through 2006, Professor Crowley has served as director for the UMR GRAVIR laboratory. Professor Crowley has edited two books, five special issues of journals, and authored over 300 articles on computer vision and mobile robotics.

Augustin Lux Augustin LUX now is professor ermeritus, his last position being Professor in Computer Science at the Ensimag (Ecole National Superieure d'Informatique et de Mathematiques Appliques) department of the Institut Polytechnique de Grenoble (Grenoble-Inp), France, where he had been teaching courses on algorithmic studies, programming, Artificial Intelligence, and machine vision. His university studies took place at the unversities of Gttingen (Vordiplom in Physik) and Grenoble (Maitrise d'Informatique), where he obtained his PhD in 1986 (Thse d'tat). Professor Lux works with the PRIMA team belonging to the LIG Laboratory and Inria, Grenoble. Research interests concern "Artificial Intelligence" in a broad sense, more particularly vision and image understanding.